ON THE STATISTICAL ANALYSIS OF FINANCIAL NETWORKS OF THE EGYPTIAN STOCK EXCHANGE

Author Detail:
Khaled Elghareni[1] -
(1)CairoUniversity, Institute Of Statistical Studies & Research,5 Ahmed Zewail St., Giza ,Egypt

Abstract: *We consider a network representation of the Egyptian Stock Market price data; this network constructed by calculating correlations between each pairs of stocks based on the closing prices over a specific time intervals .This type of networks called the "Market Graph" which make it possible to the study of the structural properties of the graph. A recent technique to summarize the stock prices data can be based on representing the stock market as a graph.*
Keywords: ***Cross-correlation, degree distribution, edge density*.**

1. INTRODUCTION

In this paper we give an overview on the statistical market data analysis based on the network approach. Based on this approach the stock prices are represented as a graph where the vertex represent stocks price and the edge represent links between stocks if the correlation coefficient with specific value met "Threshold value" . The links between stocks prices can be measured in a different ways [23]. The widely used method is to use the Pearson correlation. In the first part of this research paper different alternatives of quantifying the similarity and explain their characteristics. Some phenomena can be viewed as a complex network which show some fundamental theoretic characteristics such as epidemiology, computer networks .

While Stock markets produce every day a huge amount of data. The data are stored in different databases and data warehouses. The analysts of the stock market are looking for new and sophisticated tools for the market data analysis[16].In the context of quantifying some of complex networks in the real world appears in many applications such as ; Information networks i.e. network of citations of scientific papers, the WWW, network of collaboration of research, etc; in communications and Technological networks ,the internet, communication network, power networks, etc; Biological networks cell, neural networks, protein networks, etc;. Social and economic networks and relationship among individuals (persons, cities, firms, income of individuals, etc…) .

To model and obtain an appropriate information about such complex type of data structure; Statistics and statistical mechanics of complex networks technique are relevant to such data [22]. With the main property scale-free networks that some vertices are highly connected while most nodes of low degree and they are independent of size of nodes. The degree distribution defined by a power law relation defined by :

$P(K) \sim k^{-\gamma}$ Where the probability P(k) that the network connects with k other vertices is proportional to $k^{-\gamma}$,which give a good fit to real world data, most real networked data [12,22] exhibit a varying coefficient γ in closed interval from [2,3]

In this paper, Basic foundation of data representation of complex networks and networks topologies are represented in section two, section three is an overview about the literature of Scale-free networks with power law distributions of degrees. With special emphasis of the Economic Networks those which appear in stock markets data. Section four is the empirical findings of the characterization of the Correlation-Based Complex Networks of Modeling Stock Market and Degree-Based Index in the Egyptian Exchange. As one of the characteristics of the market graph, the correlation distribution provides Information about how the stocks are correlated to one another, thus telling us what type of market structure we are dealing with.

2. *Literature Review*

The main interest is to quantify the topological properties of the resulted complex network such as degree distribution; clustering coefficient, shortest connected path, etc… Many researchers and Data scientists observed a Long time statistical behavior of some characteristic quantities of the networks, especially those characterizing the connectedness of the networks, Degree distribution fraction of edges linked with a vertex. Clustering coefficient average number of edges linking neighborhoods of the same node, characterizing the transitivity of the networks; Length of the shortest path between two nodes; etc. Heavy tailed property in complex networks which represented by Scale-Free invariants , "Heavy tailed distributions " can be seen in many real life data e.g. income distributions; city-size distributions; Power law distributions of citations of scientific papers, of linking of web pages; etc.

[Tse and Liu 2005] constructed a complex networks to study the correlation between the stock close prices for US stocks that were traded from July 1, 2005 to August 30, 2007.the vertices are the stocks and the edges are specified by the correlation between stock price and the price returns, using a winner-take-all approach ,which determine if two vertex are connected by an edge. They found that the network of stock prices formed full information about their interdependence. And they found that the distribution of the number of links “edge “follows a power law. [Tse and Liu 2005] suggest that the variation of stock prices are strongly influenced by a moderately small number of stocks.

[YOUN,et al. 2011]used the daily prices data of 360 stocks from(Jan 1990 to August 2008) in Korean market to examine the sector dynamics of Korean Stock Market and the market volatility, and hence to explore the network structure of the Korean stock market . [YOUN et al., 2011] started with constructing a weighted network of price correlations Based on the four measures, [YOUN et al., 2011] were able to categorize "Financials," "Information Technology" and "Industrials" sectors into one group, and "Materials" and "Consumer Discretionary" sectors into another group.

[Nobi,Lee ,Lee 2014] this paper investigated How the correlation and the price- return correlation networks of 30 global indices and Korean indices were changed during the last decade ,the started with the calculation of the correlations among all the indices and choosing threshold network by assigning a threshold value and exploring the network topologies, they documented a significant change in the return correlation network topology during the global financial crises .

In line with empirical studies of large scale stock prices data sets ;[Ozsoylev,Walden:2010] used asset pricing key variables such as price and volatility as function of the network topology ,they interested in the type of networks that have a power law degree distributions, several results were obtained ,for example, the stock price volatility is a non-monotone function of network connectedness .

By using spectral and network methods in the analysis of correlation matrices of stock returns, [Heimoa and Kaskia ,2007] argued that the correlation matrices obtained from stock return time series contain information on the behavior of the market, the authors studied a subset of NYSE listed shares and applied three different techniques of analysis :spectral analysis, asset trees and asset graphs ,and by this types of analysis they illustrated which networks has a strongest correlations. [Zalesky, Fornito.,2012] demonstrates that correlation as a measure of links rise to networks with random topological properties. In particular, networks in which connectivity is measured using correlation are less clustered than non- random networks.

[Song, Dong-Ming, et al. 2012] investigate the daily correlation present among market indices of stock exchanges located all over the world in the time period Jan 1996 - Jul 2009. The main finding is that the correlation among market indices presents both a fast and a slow dynamics. They provide evidence that the short term timescale of correlation among market indices is less than 3 trading months (about 60 trading days). The average values of the non diagonal elements of the correlation matrix, correlation based graphs and the spectral properties of the largest eigenvalues and eigenvectors of the correlation matrix are carrying information about the fast and slow dynamics of correlation of market indices.

3. Graph Terminology and Notation

A graph G is a pair *(V,E)* where *V* is a set of vertices and *E* is a set of unordered pairs of elements of V (the edges). We call the order of the graph $n=|V|$ and the size of the graph $s=|E|$. We will denote the edge from v to w as vw. For $v,w\in V$ the distance *d(v,w)* is defined to be the minimum path length from v to w in E. The (closed) kth-order neighborhood (or k-neighborhood) of a vertex v is the set of vertices of distance at most k from v [6,7]:

$$N_k(v)=\{w\in V : d(v,w)\le k\}$$

The degree of a vertex v is the number of edges incident on v.

A graph: G is denoted as $G=(N,L)$, where N is the number of nodes in the network, and L is the number of links. A complete graph or full network $G=K_N$ is a graph which has a link between every pair of nodes. This graph consists of N nodes and $L=\frac{N(N-1)}{2}$ links.

Adjacent vertices: Two vertices *a* and *b* are *adjacent* if they are connected by an edge *E (a, b)*.

Adjacency matrix: Let *G (V, E)* be a graph, and |V| x |V| matrix M. An adjacent matrix M is defined as follows:

$$\begin{cases} a_{ij} = 1 \;\; if\ there\ exist\ a\ path\ from\ v_i\ to\ v_j \\ a_{ij} = 0 \;\; otherwise \end{cases}$$

Complete graph: A *complete graph* is a simple graph such that each pair of vertices is connected by an edge.

4. *A Brief Review of Power Law*

Definition (1): A r.v. X is said to have a power law distribution (in the tail) if (in the strict sense) ∃ positive constants C *and* α and x_0 such that

$$\bar{F}(x) = cx^{-\alpha} \text{ for } x \geq x_0;$$

Power Law or Scale-Free distribution is a general class of distributions with : Scale invariance feature , such feature ,when applied to a data generating process then it called scale invariant if the probability density function is similar at all scales. This power laws are widely seen and appears in different settings in biology ,physics ,demography and finance with different data generating mechanism [Gabaix et.al .2003].

scale-free networks, will have larger fluctuations from the average value. The standard method for measuring the heterogeneity of a network is to study the moments of the degree distribution.

5. *Methodology and Data*

The Egyptian Exchange, being the largest exchange in Arab world in terms of listed members has over 240 stocks listed on it. However, the top 30 stocks contribute to more than 85% of the total market capitalization of the exchange. Hence an analysis of these stocks would give us a picture of the network structure of the exchange. The timespane in this study is 10 financial years from, 1st January 2004 to 31st May 2014. Any stock which has not been in the top 30 companies consistently over these 10 years (including newly listed companies) are excluded from the analysis. This process of filtering identified 25 companies which have been in the top 30 list of the EGX over the past 10 years. The data is collected from the archives of the EGX, as published on its website.

Using this newly obtained time series of percentage changes in the prices, cross correlations are computed for every possible pair, in a time window of ±10 trading days. Cross correlation is calculated by using the relation as in [9]

$$C_{ij_T} = \frac{\sum_{t=T-m}^{t=T-1} x_t y_t}{\sqrt{\sum_{t=T-m}^{t=T-1} x_t^2 \sum_{t=T-m}^{t=T-1} y_t^2}}$$

Where x_t and y_t represent two return series.

The fundamental method for building the stock price correlation network consists of two steps;

The first step is to find the correlation between each pair of stock prices in the same time series. The second step applies a criterion to connect the stocks based on their correlation. Many algorithms to connect pairs of correlated stocks among them, and widely used method is the minimum spanning tree method. Some methods for example; planar maximally filtered graph, and winner take all method or rich get richer. The procedure for finding correlation between stocks remains the same for the above mentioned methods [11] .

Step 1: Pick the time series price data under investigation.

Step 2: For a particular time series selected from step 1, find the cross correlation for each pair of stocks using the cross correlation formula.

Step 3: Compute the cross correlation for all the stocks and create a cross correlation matrix .

Step 4: In case of the minimum spanning tree method a metric distance d_{ij} is calculated using the cross correlation matrix.[1] $d_{ij} = \left(2\left(1 - C_{ij}\right)^{0.5}\right)$

One should note that in step 4 the nodes are linked based on a threshold.

Where d_{ij} is the link or edge Euclidian distance between stock *i* and stock *j*., i.e., some high correlated nodes are discarded and low correlated nodes are retained because of the topological reduction criteria.[11] Tse, et al. introduced the winner take all connection criterion where in the drawback of minimum spanning tree and planar maximally filtered graph are eliminated.[11] In winner take all method, step 1-3 are retained.

A lose of information based on the type of algorithm to generate a topological reduction criteria ,such as minimum spanning tree and planar maximally filtered graph may occurs. For example, some strong correlated vertices may happen to be removed, or some of very weak correlated vertices remain.

Like in many Stock Market network, the degree distribution of the Egyptian Exchange market graph exhibit a power law pattern, as in Fig .1.,

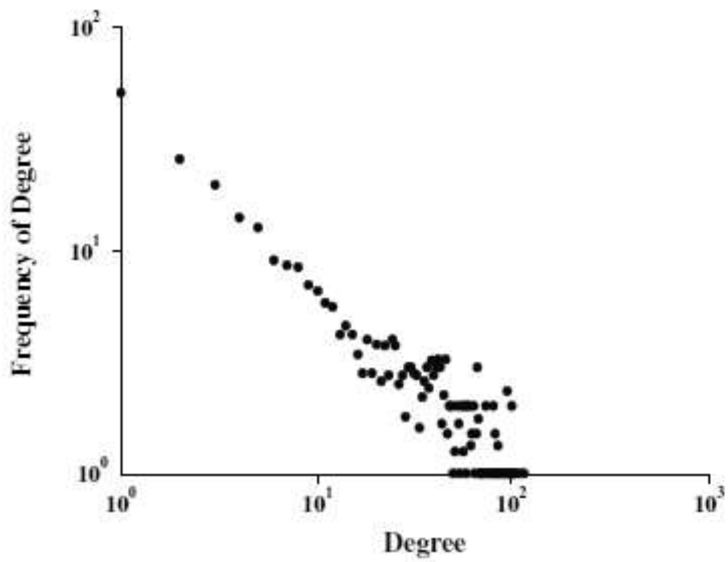
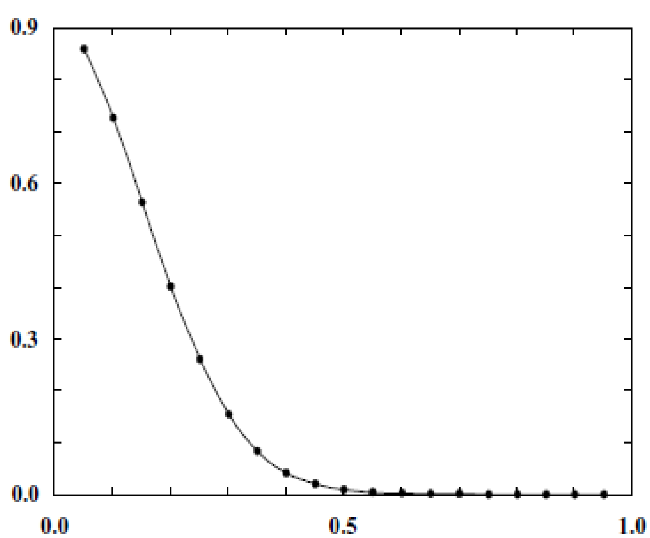


Fig.1. Frequency of degree of EGX30 networks for threshold =0.4 ,with exponent= .76.
N.B) the calculations and figures in this paper are obtained by using *R* open source software *igraph*, *Statnet*.

Edge Density



Correlation Threshold

The threshold values (θ) can be set between 0 and 1. [15] showed that for large values of threshold (0.7, 0.8, and 0.9) the stock correlation networks are scale free where the nodes linked in a manner that their degree distribution follows a power law.

Table 1. List of network parameters

Parameter	Value
Average Clustering Coefficient	0.642
Average Path Length	1.05
Diameter	2
Radius	0
Number of Weakly Connected Components	42
Average Degree	37.28

The network parameters identified (Table 1) give an idea of the network topology underlying The Egyptian Stock Exchange. clustering coefficient of the network indicates highly interacting sets of stocks in the network. The diameter of the network is small, indicating, the network is quite well connected, although the graph is not very dense.

From market graph of the largest 30 stocks traded in The Egyptian Exchange from [2003to 2013], we reviewed the cross-correlation, the frequency of degree. We found that the frequency of degree for different threshold values $\theta = 0.3, 0.4$ scales as a power law, but with different threshold values it fail to fit the power law distribution .

A. Conclusion

In this paper we dealt with the stock prices network data of the Egyptian Exchange known to be *"Market Graph Model"*. The analysis of this model provides a basically practical technique for extracting information from the stock market data. An important feature of the proposed model is the fact that it allows one to reveal certain patterns underlying the financial data; therefore, it represents a structured data mining approach. Useful information about the global properties of the stock market is noticed from the analysis of the degree distribution of the market graph. Highly specific structure of this distribution suggests that the stock market can be analyzed using the power-law model.

B. References:

[1.] Aiello, W., Chung, F. and Lu, L. 2000. A random graph model for massive

[2.] Albert, R. and Barabasi. 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47-47.

[3.] Albert, R., Jeong, H. and Barabasi, A-L. 1999. Diameter of the world wide web. Nature 401, 130-131.

[4.] Anghyuk Y., Junghoon L., and Woojin C., INT. J. Mod. Phys. C 22, 1227 (2011). Doi: 10.1142/s0129183111016853 stock market differences in correlation-based weighted network

[5.] Ashadun Nobi, Sungmin Lee, Doo Hwan Kim, Jae Woo Lee (Submitted on 7 Feb 2014)

Correlation and Network Topologies in Global and Local Stock Indices

[6.] Barabasi, A-L. and Albert, R. 1999. Emergence of Scaling in Random Networks. Science 286. 509-512.

[7.] Barabasi, A-L., Albert, R. and Jeong, H. 1999. Scale-free characteristics of random networks: the topology of the world-wide web. Physics A. 272: 173-187.

[8.] Bautin, G. A., Kalyagin, V. A., Koldanov, A. P., Koldanov, P. A., & Pardalos, P. M. (2013). Simple measure of similarity for the market graph construction.Computational Management Science, 10(2-3), 105-124.

[9.] Boginski V, Butenko S, Pardalos PM. On structural properties of the market graph. In: Nagurney A, editor. Innovations in financial and economic networks. Edward Elgar Publishers

[10.] Boginski, Vladimir, Sergiy Butenko, and Panos M. Pardalos. "Statistical analysis of financial networks." Computational statistics & data analysis 48.2 (2005): 431-443.

[11.] Chi K. Tse , Jing Liu , Francis C. M. Lau Winner-Take-All Correlation-Based Complex Networks for Modeling Stock Market and Degree-Based Indexes

[12.] Dorogovtsev, S.N., and Mendes, J.F.F. 2001. Scaling properties of scale-free, evolving networks: Continuous approach. Phys. Rev. E. 63, 056125, 1-18.

[13.] Han N. Ozsoylev, Johan Walden October 19, 2010 Asset Pricing in Large Information Networks graphs. Proceeding of the 32nd Annual ACM Symposium on Theory of Computing 171-180.

[14.] Jukka-Pekka Onnelaa,b, Kimmo Kaskiaa (2007)

Spectral and network methods in the analysis of correlation matrices of stock returns

[15.] Kleinberg, J., Kumar, S.R., Raghavan, P., Rajagopalan, S. and Tompins, A. 1999. The web as a graph: measurements, models and methods. Proc. Intl. Conf. On Combinatorics and Computing 1-18.

[16.] Koldanov, A. P., Koldanov, P. A., Kalyagin, V. A., & Pardalos, P. M. (2013). Statistical procedures for the market graph construction. Computational Statistics & Data Analysis, 68, 17-29.

[17.] Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. *Science* 296: 910-913.

[18.] Michael Mitzenmacher (2004_ Internet Mathematics Vol. 1, No. 2: 226-251 A Brief History of Generative Models for Power Law and Lognormal Distributions

[19.] [1] Newman, M. E. J., 2003, The Structure and Function of Complex Networks: SIAM Review, v. 45, p. 167-256.

[20.] [2] Newman, M. E. J., S. H. Strogatz, and D. J. Watts, 2001, Random Graphs with Arbitrary Degree Distributions and Their Applications: Phys. Rev. E, v. 64, p. 026118.

[21.] Petermann, T. and P. De Los Rios (2004). "Exploration of Scale-Free Networks." Eu. Phys.J. B 38: 201-204.

[22.] Popper, K. (1959). The Logic of Scientific Discovery. New York, Basic Books.

[23.] R. Albert; A.-L. Barabási (2002). "Statistical mechanics of complex networks".Reviews of Modern Physics 74: 47–97. arXiv:cond-mat/0106096

[24.] Song, Dong-Ming, et al. "Evolution of worldwide stock markets, correlation structure, and correlation-based graphs." Physical Review E 84.2 (2011): 026108.

[25.] Strogatz, S. 2001. Exploring random networks. Nature 410, 268-276.

[26.] Xavier Gabaix, Parameswaran Gopikrishnan, Vasiliki Plerou & H. Eugene Stanley (2003)

A theory of power-law distributions in financial market fluctuations Nature 423, 267-270 (2003)

[27.] Zalesky, A., Fornito, A., & Bullmore, E. (2012). On the use of correlation as a measure of network connectivity. Neuroimage, 60(4), 2096-2106.